Capstone: Personalized Book Recommendations Using Collaborative Filtering

Christian Armstrong

May 14, 2025

1 Problem Statement

This project addresses the development of an effective personalized book recommendation system designed to suggest books based on individual user preferences and reading history. Specifically, the system aimed to:

- Process Goodreads user-exported reading data.

- Provide accurate and personalized recommendations.

- Effectively manage both existing users and new users (cold-start problem).

- Balance recommendation diversity with relevance.

A significant challenge was handling sparse datasets and generating reliable predictions for users with minimal interaction data.

2 Initial Approach

Initially, the project implemented a hybrid recommendation system integrating:

2.1 Matrix Factorization

This involved:

- Factorizing user-item rating matrix (R) into user (U) and item (V) latent factors.

- Employing 100 latent factors initialized via Singular Value Decomposition (SVD).

- Using Adam optimizer for gradient descent optimization.

2.2 Content-Based Features

Additional content information:

- Genre data extracted from book tags.

- Representing books as genre-based feature vectors.

- Weighting these content features as 10 % of final predictions.

2.3 Popularity Metrics

Popularity-based signals incorporated:

- Number of ratings and average rating (weighted 70% and 30% respectively).

- Contribution limited to 5% of overall recommendations.

3 Challenges and Iterations

3.1 Initial Obstacles

The initial system encountered multiple challenges, notably:

- Dimension mismatches due to hybrid model complexities.

- Poor recommendations for new users (cold-start).

- Computational inefficiency from extensive feature processing.

- Difficulty balancing personalized signals with popularity and content.

3.2 Lessons Learned

Extensive experimentation revealed that:

- Complexity in hybrid methods created more noise than meaningful signals.

- Content-based features were challenging to align consistently with ratings.

- Popularity-based methods overshadowed personalized user data.

- Simplifying the recommendation approach improved accuracy significantly.

4 Refined Approach

With these insights, the project pivoted to a simpler, pure collaborative filtering model.

4.1 Collaborative Filtering Simplification

The refined approach:

- Removed extraneous content-based and popularity features.

- Retained core matrix factorization with effective SVD initialization.

- Enhanced support for new-user predictions.

4.2 Implementation Steps

The refined model follows clear steps:

1. Initialization of factors via SVD.

2. Gradient descent optimization.

3. Predictive rating computations.

4. Clamping predictions within the valid range (1–5).

4.3 Hyperparameter Optimization

Optimal hyperparameters identified through iterative experiments were:

- Latent Factors: 100

- Learning Rate: 0.003

- Epochs: 60

Performance metrics achieved included:

- RMSE: 0.846

- MAE: 0.651

Loss values demonstrated consistent and stable model improvement:

- Epoch 10: Loss = 4,505,766

- Epoch 30: Loss = 3,449,519

- Epoch 60: Loss = 1,962,647

5 Final Results

The simplified collaborative filtering approach significantly improved overall system performance:

5.1 Evaluation Metrics

- Consistent RMSE of 0.846 and MAE of 0.651.

- Efficient computation suitable for practical application.

- Stable predictions across diverse user profiles.

5.2 Quality of Recommendations

- Enhanced accuracy and relevance of recommendations.

- Improved handling of new-user scenarios.

- Increased recommendation diversity through effective tag utilization.

- Successfully managed Goodreads dataset integration.

6 Future Enhancements

Potential improvements identified include:

6.1 Technical Refinements

- Exploring advanced matrix factorization methods.

- Finding a bigger dataset.

- Optimizing computational resources for scalability.

6.2 Feature Integration

- Adding author-based or content-driven recommendations.

- Leveraging additional book metadata.

6.3 User Experience Improvements

- Providing explanations for recommendations.

- Adding interactive user feedback.

- Developing a robust web interface.

7   Conclusion

This project underscored the value of simplicity in recommendation systems. While initial hybrid solutions appeared promising, the simplified collaborative filtering model proved significantly more effective and reliable. Achieving robust performance metrics (RMSE: 0.846, MAE: 0.651), the refined approach successfully addresses the critical recommendation problem, ensuring practical applicability and scalability.